

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| | | |
|--|-----------|---|
| (51) International Patent Classification ⁵ : C07H 21/00, C12Q 1/68, C07K 15/00 | A2 | (11) International Publication Number: WO 94/25477 (43) International Publication Date: 10 November 1994 (10.11.94) |
| (21) International Application Number: PCT/IB94/00142 (22) International Filing Date: 25 April 1994 (25.04.94) (30) Priority Data: 08/054,363 26 April 1993 (26.04.93) US (71)(72) Applicants and Inventors: NIELSEN, Peter, E. [DK/DK]; Hjortevanget 509, DK-2980 Kokkedal (DK). BUCHARDT, Ole [DK/DK]; Sondergardsvej 73, DK-3500 Vaerlose (DK). EGHOLM, Michael [DK/DK]; Johnstrup Alle 3, DK-1923 Frederiksberg (DK). BERG, Rolf, H. [DK/DK]; Strandvaenget 6, DK-2960 Rungsted Kyst (DK). (74) Agent: HALLYBONE, Huw, George; CarpmacIs & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB). | | (81) Designated States: AU, BR, CA, FI, HU, JP, KR, NO, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>Without international search report and to be republished upon receipt of that report.</i> |
| (54) Title: NOVEL PEPTIDE NUCLEIC ACIDS (57) Abstract A novel class of compounds, known as peptide nucleic acids, bind complementary ssDNA and RNA strands more strongly than a corresponding DNA. The peptide nucleic acids generally comprise ligands such as naturally occurring DNA bases attached to a peptide backbone through a suitable linker. | | |

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | |
|----|--------------------------|----|---------------------------------------|----|--------------------------|
| AT | Austria | GB | United Kingdom | MR | Mauritania |
| AU | Australia | GE | Georgia | MW | Malawi |
| BB | Barbados | GN | Guinea | NE | Niger |
| BE | Belgium | GR | Greece | NL | Netherlands |
| BF | Burkina Faso | HU | Hungary | NO | Norway |
| BG | Bulgaria | IE | Ireland | NZ | New Zealand |
| BJ | Benin | IT | Italy | PL | Poland |
| BR | Brazil | JP | Japan | PT | Portugal |
| BY | Belarus | KE | Kenya | RO | Romania |
| CA | Canada | KG | Kyrgyzstan | RU | Russian Federation |
| CF | Central African Republic | KP | Democratic People's Republic of Korea | SD | Sudan |
| CG | Congo | KR | Republic of Korea | SE | Sweden |
| CH | Switzerland | KZ | Kazakhstan | SI | Slovenia |
| CI | Côte d'Ivoire | LI | Liechtenstein | SK | Slovakia |
| CM | Cameroon | LK | Sri Lanka | SN | Senegal |
| CN | China | LV | Latvia | TD | Chad |
| CS | Czechoslovakia | MC | Monaco | TG | Togo |
| CZ | Czech Republic | MD | Republic of Moldova | TJ | Tajikistan |
| DE | Germany | MG | Madagascar | TT | Trinidad and Tobago |
| DK | Denmark | ML | Mali | UA | Ukraine |
| ES | Spain | MN | Mongolia | US | United States of America |
| FI | Finland | | | UZ | Uzbekistan |
| FR | France | | | VN | Viet Nam |
| GA | Gabon | | | | |

-1-

NOVEL PEPTIDE NUCLEIC ACIDS

RELATED APPLICATION

This patent application is a continuation-in-part of application PCT EP92/01219, filed May 19, 1992. The
5 entire contents of this application, which published November 26, 1992 as WO 92/20702, is incorporated herein by reference.

FIELD OF THE INVENTION

This invention is directed to compounds that are not polynucleotides yet which bind to complementary DNA and
10 RNA strands. In particular, the invention concerns compounds wherein naturally-occurring nucleobases or other nucleobase-binding moieties are covalently bound to a polyamide backbone.

BACKGROUND OF THE INVENTION

15 Oligodeoxyribonucleotides as long as 100 base pairs (bp) are routinely synthesized by solid phase methods using commercially available, fully automatic synthesis machines. The chemical synthesis of oligoribonucleotides, however, is far less routine. Oligoribonucleotides also are much less
20 stable than oligodeoxyribonucleotides, a fact which has contributed to the more prevalent use of oligodeoxyribonucleotides in medical and biological research directed to, for example, gene therapy or the regulation of transcription or translation.

25 The function of a gene starts by transcription of its information to a messenger RNA (mRNA) which, by interaction with the ribosomal complex, directs the synthesis

- 2 -

of a protein coded for by its sequence. The synthetic process is known as translation. Translation requires the presence of various co-factors and building blocks, the amino acids, and their transfer RNAs (tRNA), all of which are
5 present in normal cells.

Transcription initiation requires specific recognition of a promoter DNA sequence by the RNA-synthesizing enzyme, RNA polymerase. In many cases in prokaryotic cells, and probably in all cases in eukaryotic
10 cells, this recognition is preceded by sequence-specific binding of a protein transcription factor to the promoter. Other proteins which bind to the promoter, but whose binding prohibits action of RNA polymerase, are known as repressors. Thus, gene activation typically is regulated positively by
15 transcription factors and negatively by repressors.

Most conventional drugs function by interaction with and modulation of one or more targeted endogenous proteins, e.g., enzymes. Such drugs, however, typically are not specific for targeted proteins but interact with other
20 proteins as well. Thus, a relatively large dose of drug must be used to effectively modulate a targeted protein. Typical daily doses of drugs are from 10^{-5} - 10^{-1} millimoles per kilogram of body weight or 10^{-3} -10 millimoles for a 100 kilogram person. If this modulation instead could be
25 effected by interaction with and inactivation of mRNA, a dramatic reduction in the necessary amount of drug necessary could likely be achieved, along with a corresponding reduction in side effects. Further reductions could be effected if such interaction could be rendered site-
30 specific. Given that a functioning gene continually produces mRNA, it would thus be even more advantageous if gene transcription could be arrested in its entirety.

Oligodeoxynucleotides offer such opportunities. For example, synthetic oligodeoxynucleotides could be used as
35 antisense probes to block and eventually lead to the breakdown of mRNA. Thus, synthetic DNA could suppress translation *in vivo*. It also may be possible to modulate the

- 3 -

genome of an animal by, for example, triple helix formation using oligonucleotides or other DNA recognizing agents. However, there are a number of drawbacks associated with triple helix formation. For example, it can only be used for
5 homopurine sequences and it requires unphysiologically high ionic strength and low pH.

Furthermore, unmodified oligonucleotides are unpractical both in the antisense approach and in the triple helix approach because they have short in vivo half-lives,
10 they are difficult to prepare in more than milligram quantities and, thus, are prohibitively costly, and they are poor cell membrane penetrators.

These problems have resulted in an extensive search for improvements and alternatives. For example, the problems
15 arising in connection with double-stranded DNA (dsDNA) recognition through triple helix formation have been diminished by a clever "switch back" chemical linking whereby a sequence of polypurine on one strand is recognized, and by "switching back", a homopurine sequence on the other strand
20 can be recognized. See, e.g., McCurdy, Moulds, and Froehler, *Nucleosides*, in press. Also, good helix formation has been obtained by using artificial bases, thereby improving binding conditions with regard to ionic strength and pH.

In order to improve half life as well as membrane
25 penetration, a large number of variations in polynucleotide backbones has been undertaken, although so far not with the desired results. These variations include the use of methylphosphonates, monothiophosphates, dithiophosphates, phosphoramidates, phosphate esters, bridged
30 phosphoroamidates, bridged phosphorothioates, bridged methylenephosphonates, dephospho internucleotide analogs with siloxane bridges, carbonate bridges, carboxymethyl ester bridges, acetamide bridges, carbamate bridges, thioether, sulfoxy, sulfono bridges, various "plastic" DNAs, α -anomeric
35 bridges, and borane derivatives.

The great majority of these backbone modifications led to decreased stability for hybrids formed between the

- 4 -

modified oligonucleotide and its complementary native oligonucleotide, as assayed by measuring T_m values. Consequently, it is generally understood in the art that backbone modifications destabilize such hybrids, i.e., result
5 in lower T_m values, and should be kept to a minimum.

OBJECTS OF THE INVENTION

It is one object of the present invention to provide compounds that bind ssDNA and RNA strands to form stable hybrids therewith.

10 It is a further object of the invention to provide compounds that bind ssDNA and RNA strands.

It is another object to provide compounds wherein naturally-occurring nucleobases or other nucleobase-binding moieties are covalently bound to a peptide backbone.

15 It is yet another object to provide compounds other than RNA that can bind one strand of a double-stranded polynucleotide, thereby displacing the other strand.

It is still another object to provide therapeutic, diagnostic, and prophylactic methods that employ such
20 compounds.

SUMMARY OF THE INVENTION

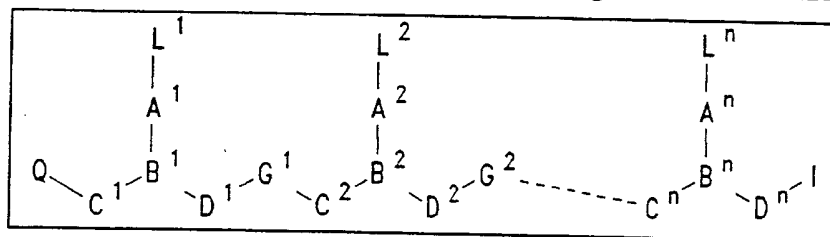
The present invention provides a novel class of compounds, known as peptide nucleic acids (PNAs), that bind complementary ssDNA and RNA strands. The compounds of the
25 invention generally comprise ligands linked to a peptide backbone. Representative ligands include either the four main naturally occurring DNA bases (i.e., thymine, cytosine, adenine or guanine) or other naturally occurring nucleobases (e.g., inosine, uracil, 5-methylcytosine or thiouracil) or
30 artificial bases (e.g., bromothymine, azaadenines or azaguanines, etc.) attached to a peptide backbone through a suitable linker.

In WO 92/20702, we described PNAs wherein such ligands are linked to a polyamide backbone solely through aza
35 nitrogen atoms. The PNAs of the invention differ from those

- 5 -

disclosed in WO 92/20702 principally in that their recognition moieties are linked to the polyamide backbone additionally through amido and/or ureido tethers.

In certain preferred embodiments, the peptide
5 nucleic acids of the invention have the general formula (I):



(I)

wherein:

n is at least 2,

each of L¹-Lⁿ is independently selected from the
10 group consisting of hydrogen, hydroxy, (C₁-C₄)alkanoyl, naturally occurring nucleobases, non-naturally occurring nucleobases, aromatic moieties, DNA intercalators, nucleobase-binding groups, heterocyclic moieties, and reporter ligands, at least one of L¹-Lⁿ being a naturally
15 occurring nucleobase, a non-naturally occurring nucleobase, a DNA intercalator, or a nucleobase-binding group;

each of C¹-Cⁿ is (CR⁶R⁷)_y where R⁶ is hydrogen and R⁷ is selected from the group consisting of the side chains of naturally occurring alpha amino acids, or R⁶ and R⁷ are
20 independently selected from the group consisting of hydrogen, (C₂-C₆)alkyl, aryl, aralkyl, heteroaryl, hydroxy, (C₁-C₆)alkoxy, (C₁-C₆)alkylthio, NR³R⁴ and SR⁵, where R³ and R⁴ are as defined above, and R⁵ is hydrogen, (C₁-C₆)alkyl, hydroxy-, alkoxy-, or alkylthio- substituted (C₁-C₆)alkyl, or R⁶ and R⁷
25 taken together complete an alicyclic or heterocyclic system;

each of D¹-Dⁿ is (CR⁶R⁷)_z where R⁶ and R⁷ are as defined above;

each of y and z is zero or an integer from 1 to 10, the sum y + z being greater than 2 but not more than 10;

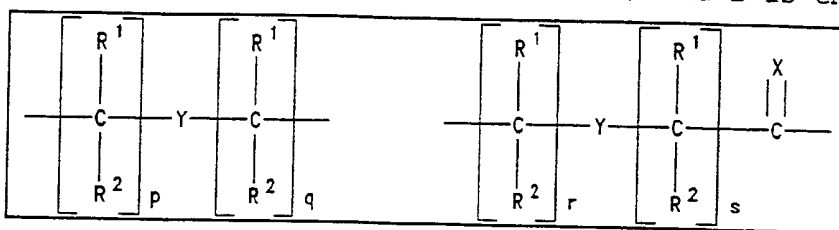
30 each of G¹-Gⁿ⁻¹ is -NR³CO-, -NR³CS-, -NR³SO- or -NR³SO₂-, in either orientation, where R³ is as defined above;

- 6 -

each pair of A^1-A^n and B^1-B^n are selected such that:

(a) A is a group of formula (IIa), (IIb) or (IIc) and B is N or R^3N^+ ; or

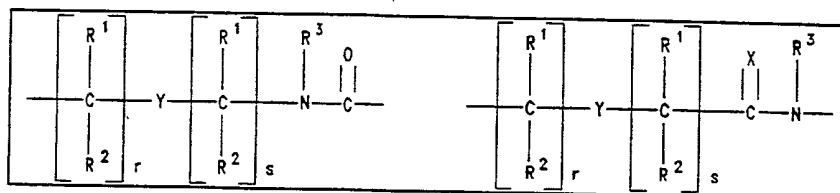
(b) A is a group of formula (IIId) and B is CH;



5

(IIa)

(IIb)



(IIc)

(IIId)

where:

X is O, S, Se, NR^3 , CH_2 or $C(CH_3)_2$;

Y is a single bond, O, S or NR^4 ;

10 each of p and q is zero or an integer from 1 to 5, the sum p+q being not more than 10;

each of r and s is zero or an integer from 1 to 5, the sum r+s being not more than 10;

15 each R^1 and R^2 is independently selected from the group consisting of hydrogen, (C_1-C_4) alkyl which may be hydroxy- or alkoxy- or alkylthio-substituted, hydroxy, alkoxy, alkylthio, amino and halogen;

20 each of G^1-G^{n-1} is $-NR^3CO-$, $-NR^3CS-$, $-NR^3SO-$ or $-NR^3SO_2-$, in either orientation, where R^3 is as defined above;

Q is $-CO_2H$, $-CONR'R''$, $-SO_3H$ or $-SO_2NR'R''$ or an activated derivative of $-CO_2H$ or $-SO_3H$; and

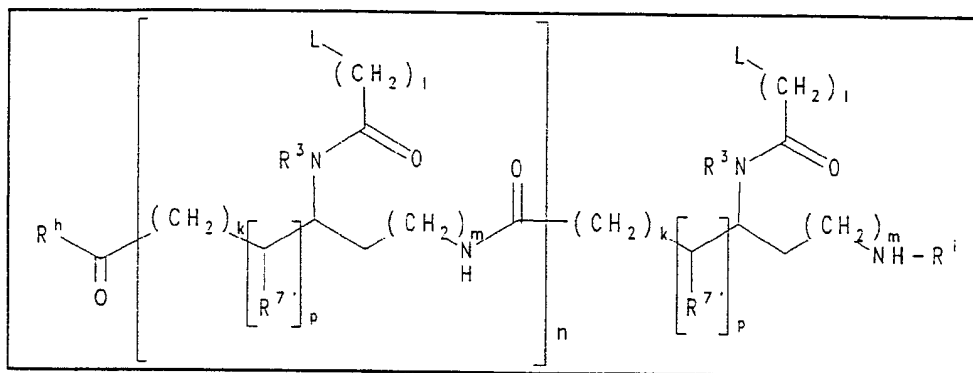
I is $-NHR'''R''''$ or $-NR'''C(O)R''''$, where R' , R'' , R''' and R'''' are independently selected from the group
25 consisting of hydrogen, alkyl, amino protecting groups,

- 7 -

reporter ligands, intercalators, chelators, peptides, proteins, carbohydrates, lipids, steroids, oligonucleotides and soluble and non-soluble polymers.

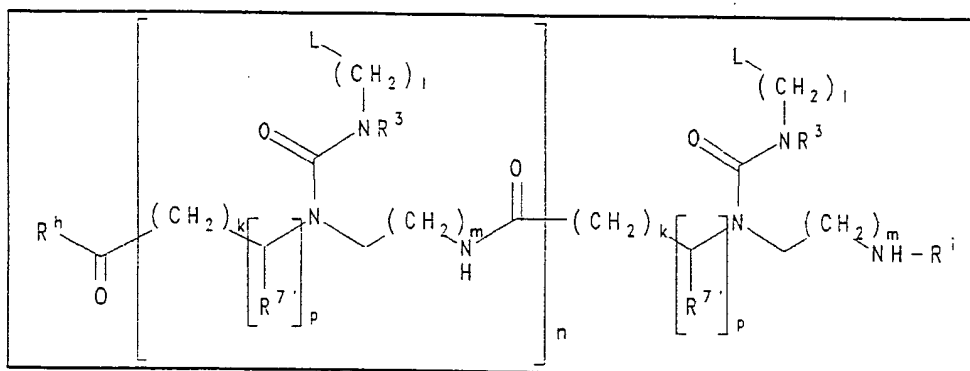
In certain embodiments, at least one A is a group of formula (IIc) and B is N or R^3N^+ . In other embodiments, A is a group of formula (IIa) or (IIb), B is N or R^3N^+ , and at least one of y or z is not 1 or 2.

Preferred peptide nucleic acids have general formula (IIIa) or (IIIb):



10

(IIIa)



(IIIb)

wherein:

each L is independently selected from the group consisting of hydrogen, phenyl, heterocyclic moieties, naturally occurring nucleobases, and non-naturally occurring nucleobases;

- 8 -

each $R^{7'}$ is independently selected from the group consisting of hydrogen and the side chains of naturally occurring alpha amino acids;

n is an integer from 1 to 60;

5 each of k , l , and m is independently zero or an integer from 1 to 5;

p is zero or 1;

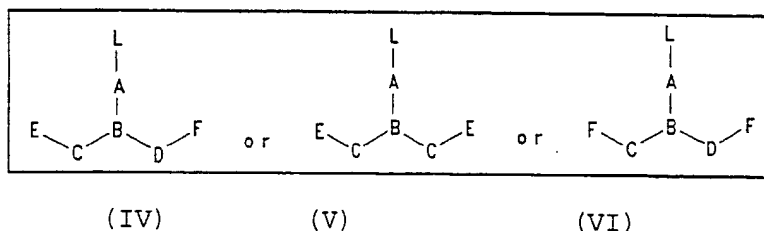
R^h is OH , NH_2 or $-NHLysNH_2$; and

R^i is H or $COCH_3$.

10 Particularly preferred are compounds having formula (IIIa) or (IIIb) wherein each L is independently selected from the group consisting of the nucleobases thymine (T), adenine (A), cytosine (C), guanine (G) and uracil (U), k and m are zero or 1, and n is an integer from 1 to 30, in particular from 4 to
15 20.

The peptide nucleic acids of the invention are synthesized by adaptation of standard peptide synthesis procedures, either in solution or on a solid phase. The synthons used are specially monomer amino acids or their
20 activated derivatives, protected by standard protecting groups. The oligonucleotide analogs also can be synthesized by using the corresponding diacids and diamines.

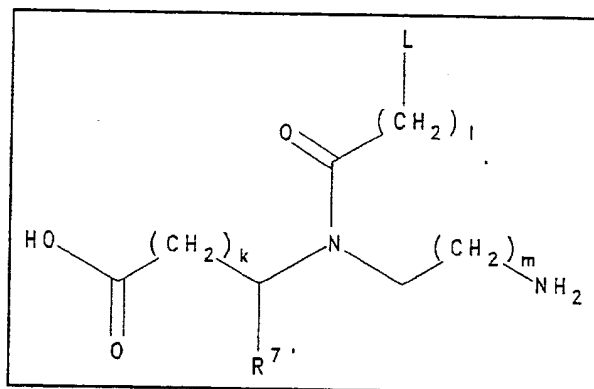
Thus, the novel monomer synthons according to the invention are selected from the group consisting of amino
25 acids, diacids and diamines having general formulae:



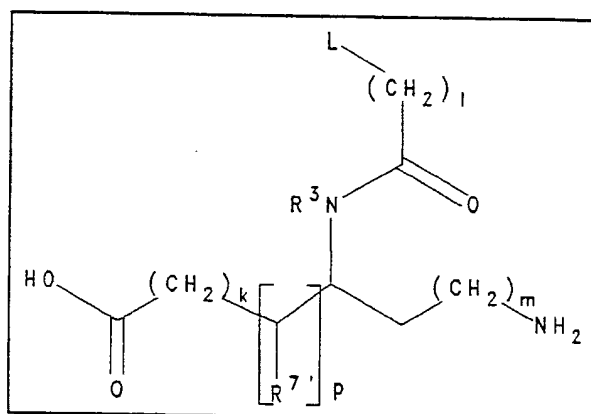
wherein L , A , E , C and D are as defined above, except that any amino groups therein may be protected by amino protecting groups; E is $COOH$, $CSOH$, $SOOH$, SO_2OH or an activated
30 derivative thereof; and F is NHR^3 or $NPgR^3$, where R^3 is as defined above and Pg is an amino protecting group.

- 9 -

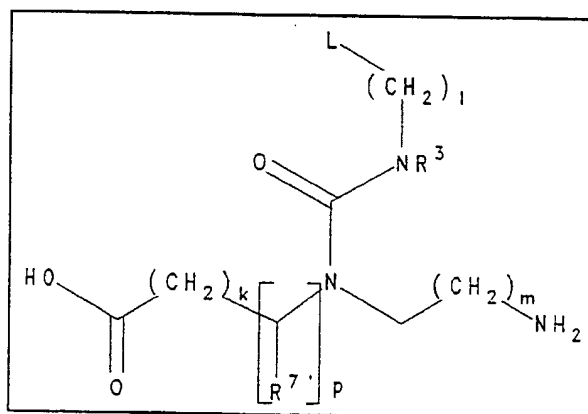
Preferred monomer synthons according to the invention have formula (VIIIa)-(VIIIc):



(VIIIa)



(VIIIb)



(VIIIc)

- 10 -

or amino-protected and/or acid terminal activated derivatives thereof, wherein L is selected from the group consisting of hydrogen, phenyl, heterocyclic moieties, naturally occurring nucleobases, and non-naturally occurring nucleobases; and R' is selected from the group consisting of hydrogen and the side chains of naturally occurring alpha amino acids.

Unexpectedly, these compounds also are able to recognize duplex DNA by displacing one strand, thereby presumably generating a double helix with the other one.

Such recognition can take place to dsDNA sequences 5-60 base pairs long. Sequences between 10 and 20 bases are of interest since this is the range within which unique DNA sequences of prokaryotes and eukaryotes are found. Reagents which recognize 17-18 bases are of particular interest since this is the length of unique sequences in the human genome.

The compounds of the invention are able to form triple helices with dsDNA and double helices with RNA or ssDNA. The compounds of the invention also are able to form triple helices wherein a first PNA strand binds with RNA or ssDNA and a second PNA strand binds with the resulting double helix or with the first PNA strand.

Whereas the improved binding of the compounds of the invention should render them efficient as antisense agents, it is expected that an extended range of related reagents may cause strand displacement, now that this surprising and unexpected new behavior of dsDNA has been discovered.

Thus, in one aspect, the present invention provides methods for inhibiting the expression of particular genes in the cells of an organism, comprising administering to said organism a reagent as defined above which binds specifically to sequences of said genes.

Further, the invention provides methods for inhibiting transcription and/or replication of particular genes or for inducing degradation of particular regions of

- 11 -

double stranded DNA in cells of an organism by administering to said organism a reagent as defined above.

Still further, the invention provides methods for killing cells or virus by contacting said cells or virus with a reagent as defined above which binds specifically to sequences of the genome of said cells or virus.

DETAILED DESCRIPTION OF THE INVENTION

In the oligonucleotide analogs and monomer synthons according to the invention, ligand L is primarily a naturally occurring nucleobase attached at the position found in nature, i.e., position 9 for adenine or guanine, and position 1 for thymine or cytosine. Alternatively, L may be a non-naturally occurring nucleobase (nucleobase analog), another base-binding moiety, an aromatic moiety, (C₁-C₄)alkanoyl, hydroxy or even hydrogen. It will be understood that the term nucleobase includes nucleobases bearing removable protecting groups. Some typical nucleobase ligands and illustrative synthetic ligands are shown in Figure 2 of WO 92/20702. Furthermore, L can be a DNA intercalator, a reporter ligand such as, for example, a fluorophor, radio label, spin label, hapten, or a protein-recognizing ligand such as biotin. In monomer synthons, L may be blocked with protecting groups, as illustrated in Figure 4 of WO 92/20702.

Linker A can be a wide variety of groups such as -CR¹R²CO-, -CR¹R²CS-, -CR¹R²CSe-, -CR¹R²CNHR³-, -CR¹R²C=CH₂- and -CR¹R²C=C(CH₃)₂-, where R¹, R² and R³ are as defined above. Preferably, A is methylenecarbonyl (-CH₂CO-), amido (-CONR³-), or ureido (-NR³CONR³-). Also, A can be a longer chain moiety such as propanoyl, butanoyl or pentanoyl, or corresponding derivative, wherein O is replaced by another value of X or the chain is substituted with R¹R² or is heterogenous, containing Y. Further, A can be a (C₂-C₆)alkylene chain, a (C₂-C₆)alkylene chain substituted with

- 12 -

R^1R^2 or can be heterogenous, containing Y. In certain cases, A can just be a single bond.

In one preferred form of the invention, B is a nitrogen atom, thereby presenting the possibility of an achiral backbone. B can also be R^3N^+ , where R^3 is as defined above, or CH.

In the preferred form of the invention, C is $-CR^6R^7-$, but can also be a two carbon unit, i.e. $-CHR^6CHR^7-$ or $-CR^6R^7CH_2-$, where R^6 and R^7 are as defined above. R^6 and R^7 also can be a heteroaryl group such as, for example, pyrrolyl, furyl, thienyl, imidazolyl, pyridyl, pyrimidinyl, indolyl, or can be taken together to complete an alicyclic system such as, for example, 1,2-cyclobutanediyl, 1,2-cyclopentanediy l or 1,2-cyclohexanediy l.

In the preferred form of the invention, E in the monomer synthon is COOH or an activated derivative thereof, and G in the oligomer is $-CONR^3-$. As defined above, E may also be CSOH, SOOH, SO_2OH or an activated derivative thereof, whereby G in the oligomer becomes $-CSNR^3-$, $-SONR^3-$ and $-SO_2NR^3-$, respectively. The activation may, for example, be achieved using an acid anhydride or an active ester derivative, wherein hydrogen in the groups represented by E is replaced by a leaving group suited for generating the growing backbone.

The amino acids which form the backbone may be identical or different. We have found that those based on 2-aminoethylglycine are especially well suited to the purpose of the invention.

In some cases it may be of interest to attach ligands at either terminus (Q, I) to modulate the binding characteristics of the PNAs. Representative ligands include DNA intercalators which will improve dsDNA binding or basic groups, such as lysine or polylysine, which will strengthen the binding of PNA due to electrostatic interaction. To decrease negatively charged groups such as carboxy and sulfo

- 13 -

groups could be used. The design of the synthons further allows such other moieties to be located on non-terminal positions.

In a further aspect of the invention, the PNA
5 oligomers are conjugated to low molecular effector ligands such as ligands having nuclease activity or alkylating activity or reporter ligands (fluorescent, spin labels, radioactive, protein recognition ligands, for example, biotin or haptens). In a further aspect of the invention, the PNAs
10 are conjugated to peptides or proteins, where the peptides have signaling activity and the proteins are, for example, enzymes, transcription factors or antibodies. Also, the PNAs can be attached to water-soluble or water-insoluble polymers. In another aspect of the invention, the PNAs are conjugated
15 to oligonucleotides or carbohydrates. When warranted, a PNA oligomer can be synthesized onto some moiety (e.g., a peptide chain, reporter, intercalator or other type of ligand-containing group) attached to a solid support.

Such conjugates can be used for gene modulation
20 (e.g., gene targeted drugs), for diagnostics, for biotechnology, and for scientific purposes.

As a further aspect of the invention, PNAs can be used to target RNA and ssDNA to produce both antisense-type gene regulating moieties and hybridization probes for the
25 identification and purification of nucleic acids. Furthermore, the PNAs can be modified in such a way that they can form triple helices with dsDNA. Reagents that bind sequence-specifically to dsDNA have applications as gene targeted drugs. These are foreseen as extremely useful drugs
30 for treating diseases like cancer, AIDS and other virus infections, and may also prove effective for treatment of some genetic diseases. Furthermore, these reagents may be used for research and in diagnostics for detection and isolation of specific nucleic acids.

- 14 -

The triple helix principle is believed to be the only known principle in the art for sequence-specific recognition of dsDNA. However, triple helix formation is largely limited to recognition of homopurine-homopyrimidine sequences. Strand displacement is superior to triple helix recognition in that it allows for recognition of any sequence by use of the four natural bases. Also, in strand displacement recognition readily occurs at physiological conditions, that is, neutral pH, ambient (20-40 C) temperature and medium (100-150 mM) ionic strength.

Gene targeted drugs are designed with a nucleobase sequence (containing 10-20 units) complementary to the regulatory region (the promoter) of the target gene. Therefore, upon administration of the drug, it binds to the promoter and block access thereto by RNA polymerase. Consequently, no mRNA, and thus no gene product (protein), is produced. If the target is within a vital gene for a virus, no viable virus particles will be produced. Alternatively, the target could be downstream from the promoter, causing the RNA polymerase to terminate at this position, thus forming a truncated mRNA/protein which is nonfunctional.

Sequence-specific recognition of ssDNA by base complementary hybridization can likewise be exploited to target specific genes and viruses. In this case, the target sequence is contained in the mRNA such that binding of the drug to the target hinders the action of ribosomes and, consequently, translation of the mRNA into protein. The peptide nucleic acids of the invention are superior to prior reagents in that they have significantly higher affinity for complementary ssDNA. Also, they possess no charge and water soluble, which should facilitate cellular uptake, and they contain amides of non-biological amino acids, which should make them biostable and resistant to enzymatic degradation by, for example, proteases.

- 15 -

It is believed that PNA oligomers according to the invention exhibit biochemical/biological properties similar to those disclosed in WO 92/20702, and that such properties can be determined by similar means. It also is believed that the PNAs of the invention can be synthesized by similar methodology. Monomer synthons according to the invention are coupled using the standard protocols to give the desired oligomeric sequences.

One monomer synthon according to the invention is prepared by reacting glycine hydrochloride 1 with ethyl acrylate in the presence of an acid scavenging base to give the Michael adduct, N-carboxamidomethyl- β -alanine ethyl ester 2. The adduct 2 is condensed with 1-carboxymethyl thymine 3 using diisopropylcarbodiimide and hydroxybenzotriazole to give (N-carboxamidomethyl)-N-(1-(thymine-1-yl)acetyl)- β -alanine ethyl ester 4. The primary amide of 4 is oxidized and rearranged to the Boc-protected amine with sodium hypobromite in t-butanol to provide (N-t-butyloxycarbonylaminoethyl)-N-(1-(thymine-1-yl)acetyl)- β -alanine ethyl ester 5. The ethyl ester is hydrolyzed with aqueous base to provide the thymine-based monomer, (N-t-butyloxycarbonylaminoethyl)-N-(1-(thymine-1-yl)acetyl)- β -alanine 6. This reaction sequence is followed to prepare the corresponding C, G, and A -based monomers, namely, N-(t-butyloxycarbonylaminoethyl)-N-(1-(N⁴-benzyloxycarbonylcytosine-1-yl)acetyl)- β -alanine, N-(t-butyloxycarbonylaminoethyl)-N-(1-(2-amino-6-benzyloxy-purine-9-yl)acetyl)- β -alanine, N-(t-butyloxycarbonylaminoethyl)-N-(1-(N⁶-benzyloxycarbonyl-adenine-9-yl)acetyl)- β -alanine.

A further monomer synthon is prepared by reacting 1-aminothymine with triphosgene to give the carbamoyl chloride derivative, 8, which is condensed with N-(2-t-butyloxycarbonylaminoethyl)glycine ethyl ester and an acid scavenger to yield the fully protected monomer, 9. The ester is hydrolyzed to give the useful monomer, 10. This reaction

- 16 -

sequence is followed to prepare the corresponding C, G, and A-based monomers, namely, N-(t-butyloxycarbonylaminoethyl)-N-(1-N⁴-benzyloxycarbonyl-cytosin-1-yl)aminocarbonyl)-glycine, N-(t-butyloxycarbonylaminoethyl)-N-(1-(2-amino-6-benzyloxy-purin-9-yl)-aminocarbonyl)-glycine, N-(t-butyloxycarbonyl-amino ethyl)-N-(1-(N⁶-benzyloxycarbonyl-adenine-9-yl)aminocarbonyl)-glycine.

A further monomer synthon is prepared by converting 2-hydroxy-5-(t-butyloxycarbonylamino)pentanoic acid ethyl ester to its azido analog via the use of diphenyl phosphoryl azide, DEAD, and triphenylphosphine generally by the procedure described in *Tetrahedron Letters*, (1977), p. 1977. The azido compound, 12, was converted to the iminophosphorane, 13, and used immediately in a high pressure reaction with carbon dioxide to convert it into isocyanate, 14. The isocyanate is condensed with thymine to give the fully protected monomer, 15, which is hydrolyzed to the actual monomer, 16, using hydroxide. This reaction sequence is followed to prepare the corresponding C, G, and A-based monomers, namely, 5-(t-butyloxycarbonylamino)-2-((N⁴-benzyloxycarbonyl-cytosin-1-yl)carbonylamino-pentanoic acid ethyl ester, 5-(t-butyloxycarbonylamino)-2-((2-amino-6-benzyloxy-purin-9-yl)carbonylamino)-pentanoic acid ethyl ester, 5-(t-butyloxycarbonylamino)-2-((N⁶-benzyloxycarbonyl-adenine-9-yl)carbonylamino)-pentanoic acid ethyl ester.

Additional objects, advantages, and novel features of this invention will become apparent to those skilled in the art upon examination of the following examples thereof, which are not intended to be limiting.

30 Example 1

N-carboxamidomethyl- β -alanine ethyl ester, 2.

Glycinamide hydrochloride (1, 11.0 g, 0.10 mol) is suspended in 500 mL of dioxane and diisopropylethylamine (12.9 g, 0.10 mol) is added and the mixture cooled to 0°C.

- 17 -

With stirring ethyl acrylate (10.0 g, 0.10 mol) is added dropwise over 15 minutes. After the addition is complete the reaction is allowed to warm to room temperature and stir for 12 hours. The reaction mixture is diluted with water 1.5 L and the pH adjusted to 4. The solution is extracted with diethyl ether (3X300 mL). The aqueous layer is neutralized with sodium hydroxide and extracted 5 times with dichloromethane. The dichloromethane extracts are combined, dried (Na_2SO_4), and the solvent removed to give a solid.

10 Example 2

(N-carboxamidomethyl)-N-(1-(thymine-1-yl)acetyl)- β -alanine ethyl ester, 4.

The product from Example 1, 2, is dissolved in dichloromethane (500 mL) and to this is added 1-carboxymethyl thymine (3, 15.5 g, 0.1 mol), hydroxybenzotriazole (13.5 g, 0.1 mol) and the solution is cooled to 0°C in an ice bath. Diisopropylcarbodiimide (12.6 g, 0.1 mol) dissolved in 50 mL of dichloromethane is added in one portion and the reaction is stirred for 12 hours. The suspended solids are removed by filtration and washed with dichloromethane. The solution is evaporated to a solid and the desired product, 4, is obtained after chromatography on silica gel using dichloromethane/ethanol as eluent.

Example 3

25 (N-t-butyloxycarbonylaminoethyl)-N-(1-(thymine-1-yl)acetyl)- β -alanine ethyl ester, 5.

The product from Example 2, 4, is dissolved t-butanol/dioxane (4:1, 500 mL), cooled to 0°C, and sodium hypobromite solution (0.15 mol) is added. After 6 hours the reaction mixture is evaporated to remove volatile solvents and the residue is diluted with water (500 mL) and extracted (5X200 mL) with dichloromethane. The extracts are combined, dried, and evaporated to a solid.

- 18 -

Example 4

(N-t-butyloxycarbonylaminomethyl)-N-(1-(thymine-1-yl)acetyl)- β -alanine, 6.

The product from Example 3, 5, is dissolved in
5 ethanol (500 mL) and 5M sodium hydroxide (20 mL) is added.
The solution is stirred for 6 hours, then neutralized with 5N
hydrochloric acid (20 mL) and the solution evaporated to a
solid. This solid is recrystallized to give the title
compound.

10 Example 5

1-(Chlorocarbonylamino)-thymine, 8.

1-Amino thymine (7, 12.5 g, 0.1 mol) is dissolved
in tetrahydrofuran (500 mL) and the solution is cooled to 0°C
and a 2M solution of triphosgene in THF (150 mL) is added and
15 the reaction is stirred for 4 hours. The solution is
evaporated to a solid, which is used as is in the next
reaction.

Example 6

N-(2-t-Butyloxycarbonylaminoethyl)-N-(thymine-1-yl-amino-
20 carbonyl)glycine ethyl ester, 9.

The product from Example 5, 8, is dissolved in THF
(500 mL) and diisopropylethylamine (12.9 g, 0.1 mol) is
added, followed by N-(2-t-butyloxycarbonylaminoethyl)glycine
ethyl ester (24.6 g, 0.1 mol) and the solution stirred for 12
25 hours. The reaction is diluted with 1000 mL of diethyl ether
and extracted 3 times with 0.1N HCl solution. The organic
layer is washed with diluted sodium bicarbonate solution,
dried, filtered and evaporated to give a solid.

Example 7

30 N-(2-t-Butyloxycarbonylaminoethyl)-N-(thymine-1-yl-amino-
carbonyl)glycine, 10.

The product from Example 6, 9, is dissolved in
ethanol (500 mL) and 2M sodium hydroxide (50 mL) is added.
The reaction is stirred for 6 hours, then neutralized with 50

- 19 -

mL of 2M HCl solution, and evaporated to remove the ethanol. The residue is dissolved in dichloromethane (250 mL) and is extracted with water (2X50 mL), dried, filtered, and evaporated to a solid.

5 **Example 8**

2-Azido-5-(t-butyloxycarbonylamino)pentanoic acid ethyl ester, 12.

2-Hydroxy-5-(t-butyloxycarbonylamino)pentanoic acid ethyl ester, (11, 26.1 g, 0.1 mol), triphenylphosphine (26.2 g, 0.10 mol), diethylazodicarboxylate (17.4 g, 0.1 mol), and diphenylphosphorylazide (27.5 g, 0.1 mol) is dissolved in THF (500 mL) and heated to reflux and maintained there for 8 hours. The reaction is cooled to room temperature, evaporated to an oil, and the product isolated by column
15 chromatography using dichloromethane:ethanol as eluent.

Example 9

2-Iminotriphenylphosphoranyl-5-(t-butyloxycarbonylamino)pentanoic acid ethyl ester, 13.

The product, 12, from Example 8 is dissolved in THF
20 and triphenylphosphine (26.2 g, 0.1 mol) is added and the reaction is stirred for 4 hours. This solution is used as is for the next reaction (Example 10).

Example 10

**2-Isocyanato-5-(t-butyloxycarbonylamino)pentanoic acid ethyl
25 ester, 14.**

The reaction solution from Example 9 is placed in a Parr® bomb and carbon dioxide (22 g, 0.5 mol) is condensed into the bomb. The bomb is sealed and heated to 50°C for 12 hours. The bomb is cooled and vented to atmospheric
30 pressure. The solution is transferred from the bomb to a flask and used as is in the next reaction (Example 11).

- 20 -

Example 11

2-(Thymin-1-ylcarbonylamino)-5-(t-butylloxycarbonylamino)pentanoic acid ethyl ester, 15.

The reaction solution from Example 10 is placed in a flask and to this is added thymine (12.6 g, 0.1 mol). The resulting solution is allowed to stir for 12 hours, then is evaporated to a solid, which is purified by column chromatography using dichlormethane:ethanol as the eluent.

Example 12

2-(Thymin-1-ylcarbonylamino)-5-(t-butylloxycarbonylamino)pentanoic acid, 16.

The product from Example 11, 15, is dissolved in ethanol (500 mL) and to this added 2M sodium hydroxide (50 mL) and the reaction stirred for 12 hours. The reaction is neutralized with 2M HCl solution (50 mL) and evaporated to a small volume. This residue is diluted with water (250 mL) and extracted with dichloromethane (4X100 mL), dried, filtered, and evaporated to give a solid.

Example 13

1-(2(-Thyminyl)acetyl)-1-(2-(tBoc-aminopropyl))glycine, 17

1,3-Diaminopropane (0.05 mmol) was dissolved in THF (100 mL) and chloroacetic acid (0.045 mmol) was added and the reaction heated at reflux for 4 hours and cooled to room temperature. The solution was diluted with diethyl ether (500 mL) and extracted 3 times with 1N NaOH solution. The combined water layers were acidified to pH = 4 and extracted with dichloromethane (5X50 mL). The organic layers were combined, dried, filtered and evaporated to an oil. This oil was dissolved in methanol (1000 mL) and dry HCl gas added. The reaction was heated to reflux and maintained there for 8 hours. The reaction was cooled and evaporated to an oil. This oil was dissolved in dioxane/water and p-nitrophenyl-t-butylcarbonate (0.05 mmol) was added and the pH adjusted to 10. The reaction was stirred for 4 hours, then neutralized and extracted 5 times with dichloromethane. The methyl ester

- 21 -

was dissolved in 50% DMF in dichloromethane and to this was added dicyclohexylcarbodiimide (DCC, 0.05 mmol) and hydroxbenzotriazole (0.05 mmol), and 2-thyminylacetic acid (0.05 mmol). The reaction was stirred for 18 hours then the DCC was removed by filtration and the residue evaporated to an oil. The oil was purified by column chromatography.

Example 14**3-(Boc-amino)-1,2-propanediol, 18**

3-Amino-1,2-propanediol (40.00 g, 0.440 mol, 1.0 eqv) was dissolved in water (1000 ml) and cooled to 0 °C, and di-tert-butyl dicarbonate (115.0 g, 0.526 mol, 1.2 eqv) was added in one portion. The reaction mixture was heated to room temperature on a water bath with stirring. The pH was maintained at 10.5 with a solution of sodium hydroxide (17.56 g, 0.440 mol, 1.0 eqv) in water (120 ml). When the addition of aqueous sodium hydroxide was completed, the reaction mixture was stirred overnight at room temperature. Subsequently, ethyl acetate (750 ml) was added to the reaction mixture followed by cooling to 0°C and the pH was adjusted to 2.5 with 4N sulfuric acid with vigorous stirring. The phases were separated. The water phase was washed with additional ethyl acetate (6x350 ml). The volume of the organic phase was reduced to 900 ml by evaporation under reduced pressure and washed with a saturated aqueous solution of potassium hydrogen sulfate diluted to twice its volume (1x1000 ml) and with saturated aqueous sodium chloride (1x500 ml). The organic phase was dried (MgSO₄) and evaporated under reduced pressure to yield 50.12 g (60%) of the title compound. The product could be solidified by evaporation from methylene chloride and subsequent freezing. ¹H-NMR (CDCl₃/TMS): δ = 1.43 (s, 9H, Me₃C), 3.25 (m, 2H, CH₂), 3.57 (m, 2H, CH₂), 3.73 (m, 1H, CH). ¹³C-NMR (CDCl₃/TMS): δ = 28.2 (Me₃C), 42.6 (CH₂), 63.5, 71.1 (CH₂OH, CHOH), 79.5 (Me₃C), 157.0 (C=O).

- 22 -

Example 15**Boc-aminoacetaldehyde, 19**

3-(Boc-amino)-1,2-propanediol (18, 20.76 g, 0.109 mol, 1 eqv) was suspended in water (150 ml). Potassium m-
5 periodate (24.97 g, 0.109 mol, 1 eqv) was added and the reaction mixture was stirred for 2 h at room temperature under nitrogen. The reaction mixture was filtered and the water phase was extracted with chloroform (6x250 ml). The organic phase was dried (MgSO₄) and evaporated to afford
10 crude Boc-aminoacetaldehyde as a golden oil. This oil was kugelrohr distilled at 80°C and 0.2 mbar to yield 13.19 g (76%) of the title compound as a semicrystalline solid. ¹H-NMR (DMSO-d₆/TMS): δ = 1.47 (s, 9H, Me₃C), 3.81 (d, J=5.6 Hz, 2H, CH₂), 7.22 (b, 1 H, NH), 9.54 (s, 1 H, CHO). ¹³C-NMR (DMSO-d₆/TMS): δ = 28.2 (Me₃C), 50.5 (CH₂), 78.4 (Me₃C), 156.1
15 d₆/TMS): δ = 28.2 (Me₃C), 50.5 (CH₂), 78.4 (Me₃C), 156.1 (carbamate C=O), 200.6 (CHO). Anal. Calcd. for C₇H₁₃NO₃: C, 52.82; H, 8.23; N, 8.80. Found: C, 52.21; H, 8.15; N, 8.46.

Example 16**(Boc-amino)ethylglycine Methyl Ester, 20****20 A. Reduction With Sodium Cyanoborohydride**

Boc-aminoacetaldehyde (19, 1.00 g, 6.3 mmol, 1 eqv) was dissolved in methanol (50 ml). Anhydrous sodium acetate (1.03 g, 12.6 mmol, 2 eqv), glycine methyl ester hydrochloride (Aldrich Chemical Co., 0.79 g, 6.3 mmol, 1 eqv)
25 and sodium cyanoborohydride (1.97 g, 31.4 mmol, 5 eqv) were added to the solution in that order. The reaction mixture was stirred for 2 h at room temperature under nitrogen. Water (50 ml) was added to the suspension and the resulting clear solution was evaporated under reduced pressure to
30 remove the methanol. The aqueous phase was extracted with methylene chloride (3x100 ml). The organic phase was washed with a saturated aqueous solution of sodium chloride (1x100 ml), dried (Na₂SO₄), filtered and then evaporated under reduced pressure affording 1.41 g of crude title compound as
35 a yellow oil. The crude product was kugelrohr distilled at 110°C and 0.5 mbar to yield 0.49 g (34%) of 2-(Boc-

- 23 -

amino)ethylglycine methyl ester as a colorless liquid. ^1H -NMR (CDCl_3/TMS): δ = 1.36 (s, 9H, Me_3C), 1.91 (s, 1H, NH), 2.67 (t, $J=6$ Hz, 2H, NHCH_2), 3.13 (q, $J=6$ Hz, 2H, NHCH_2), 3.34 (s, 2H, CH_2COO), 3.65 (s, 3H, OMe), 5.13 (b, 1H, carbamate NH). ^{13}C -NMR (CDCl_3/TMS): δ = 28.2 (Me_3C), 39.9, 48.5 (NHCH_2), 50.0 (CH_2COO), 51.5 (OMe), 78.9 (Me_3C), 155.9 (carbamate C=O), 172.6 (ester C=O). Anal. Calcd for $\text{C}_{10}\text{H}_{12}\text{N}_2\text{O}_4$: C, 51.71; H, 8.68; N, 12.06. Found: C, 51.55; H, 8.72; N, 11.79.

10 B. Catalytic hydrogenation

Boc-aminoacetaldehyde (2.08 g, 13.1 mmol, 1 eqv) was dissolved in methanol (50 ml) and cooled to 0 °C. Palladium on activated carbon (10%, 0.4 g) was added under nitrogen and with vigorous stirring. Anhydrous sodium acetate (2.14 g, 26.1 mmol, 2 eqv) and glycine methyl ester, hydrochloride (1.64 g, 13.1 mmol, 1 eqv) each dissolved in methanol (25 ml) were added to the mixture. The reaction mixture was hydrogenated at atmospheric pressure and room temperature with vigorous stirring, until hydrogen uptake had ceased (when 287 ml, 13.1 mmol, 1 eqv had been consumed) after about 1 h. The reaction mixture was filtered and the solvent removed under reduced pressure. The residue was suspended in water (30 ml), and during vigorous stirring pH was adjusted to 8 by dropwise addition of 0.5 N NaOH. The water phase was extracted with methylene chloride (4x50 ml). The organic phase was dried (Na_2SO_4), filtered and evaporated under reduced pressure to yield 3.03 g of crude title compound as a golden oil. The crude product was kugelrohr distilled at 100 °C and 0.2 mbar to afford 2.33 g (77%) of 2-(Boc-amino)ethylglycine methyl ester as a colorless liquid. The analytical data were in accord with those provided above for the reduction with sodium cyanoborohydride.

Example 17**General Method for the Synthesis of PNA Oligomers**

Oligomers were prepared generally in accordance with the methods disclosed by WO 92/20702. Benzyhydrylamine resin (initially loaded 0.28 mmol/gm with Boc-L-Lys(2-chlorobenzyloxycarbonyl)) was swollen in DMF and an excess of a monomer to be coupled was added, followed by dicyclohexylcarbodiimide (0.15M in 50% DMF in dichloromethane). The Boc deprotection was accomplished by trifluoroacetic acid treatment. The progress of the coupling reactions was monitored by quantitative ninhydrin analysis. The PNA was released from the resin using anhydrous HF under standard conditions. The products were purified using HPLC with acetonitrile-water (0.1%TFA) gradient and structure confirmed by fast atom bombardment mass spectrometry. The following sequences have been synthesized by this method:

H-T₁₀LysNH₂
H-T₄CT₅LysNH₂
H-T₂CT₂CT₄LysNH₂
20 H-T₄CT₂CT₂LysNH₂
H-TGTACGTCACAACTA-NH₂
H-CCTTCCCTT-NH₂
H-TTCCCTTCC-NH₂
H-TAGTTATCTCTATCT-NH₂
25 H-TGTACGTCACAACTA-NH₂
H-GCACAGCC-LYS-NH₂
H-TTTTCTTTT-NH₂
H-TTTTTTTTTTCCCCCCC-NH₂
H-CCCCCCTTTTTTTTTT-NH₂
30 H-CCTCCTTCCC-NH₂
H-TTCTCTCTCT-NH₂
H-TTTTTCTCTCTCTCT-NH₂
H-CCCCCACCCTTCCCCCTCTC-(Lys)₃NH₂
H-CTTATATTCCGTCATCGCTCLys-NH₂
35 H-CTGTCTCCATCCTCTTCACT-NH₂
H-TATTCGTCATCGCTCCTCALys-NH₂

- 25 -

H-CCCCCACCACCTTCCCCCTCTC-NH₂H-CTGCTGCCTCTGTCTCAGGTLysNH₂H-T₄-(β-alanine)C-T₅LysNH₂H-T₄-(β-alanine)T-T₅LysNH₂

- 5 Those skilled in the art will appreciate that numerous changes and modifications may be made to the preferred embodiments of the invention and that such changes and modifications may be made without departing from the spirit of the invention. It is therefore intended that the
- 10 appended claims cover all such equivalent variations as fall within the true spirit and scope of the invention.

- 26 -

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Nielsen, Peter E.
Buchardt, Ole
Egholm, Michael
Berg, Rolf H.
- (ii) TITLE OF INVENTION: Novel Peptide Nucleic Acids
- (iii) NUMBER OF SEQUENCES: 24
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz, and
Norris
 - (B) STREET: One Liberty Place - 46th Floor
 - (C) CITY: Philadelphia
 - (D) STATE: PA
 - (E) COUNTRY: USA
 - (F) ZIP: 19103
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Lucci, Joseph
 - (B) REGISTRATION NUMBER: 33,307
 - (C) REFERENCE/DOCKET NUMBER: ISIS1017
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 215-568-3100
 - (B) TELEFAX: 215-568-3439

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 11 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: peptide
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 1
 - (D) OTHER INFORMATION: /label= MODIFIED-SITE
/note= "Thymine heterocycle is attached to

- 27 -

N-acetyl(2-aminoethyl)glycine through the
N-acetyl group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 2
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 3
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 4
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 6
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- 28 -

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 8
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 9
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 10
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Lys
1 5 10

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 2
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

- 29 -

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 3
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 4
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 6
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 8
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 9

- 30 -

- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 10
(D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Lys
1 5 10
- (2) INFORMATION FOR SEQ ID NO:3:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown
(ii) MOLECULE TYPE: peptide
(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 1
(D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 2
(D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 3
(D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
(A) NAME/KEY: Modified-site

- 31 -

- (B) LOCATION: 4
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 5
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 6
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 7
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 8
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 9
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 10
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to

- 32 -

N-acetyl(2-aminoethyl)glycine through the N-acetyl group at position 1 of the heterocycle."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Lys
1 5 10

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 2
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 3
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 4
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /label= Modified-site

- 33 -

/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 6

(D) OTHER INFORMATION: /label= Modified-site

/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 7

(D) OTHER INFORMATION: /label= Modified-site

/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 8

(D) OTHER INFORMATION: /label= Modified-site

/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 9

(D) OTHER INFORMATION: /label= Modified-site

/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 10

(D) OTHER INFORMATION: /label= Modified-site

/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Lys
1 5 10

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15 amino acids

- 34 -

- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: peptide
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 1
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 2
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Guanine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 3
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 4
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 5
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 6
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Guanine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl

- 35 -

group at position 9 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 8
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 9
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 10
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 11
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 12
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site

- 36 -

(B) LOCATION: 13

(D) OTHER INFORMATION: /label= Modified-site

/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 14

(D) OTHER INFORMATION: /label= Modified-site

/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 15

(D) OTHER INFORMATION: /label= Modified-site

/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa |
| 1 | | | | 5 | | | | | 10 | | | | | 15 |

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 1

(D) OTHER INFORMATION: /label= Modified-site

/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 2

(D) OTHER INFORMATION: /label= Modified-site

/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

- 37 -

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 3
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 4
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 6
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 8
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 9

- 38 -

(D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1 5

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 2
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 3
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 4
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

- 39 -

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 6
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 8
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 9
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1 5

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single

- 40 -

- (D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: peptide
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 1
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 2
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 3
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Guanine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 4
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 5
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 6
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."
- (ix) FEATURE:

- 41 -

- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 7
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 8
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 9
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 10
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 11
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 12
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 13
 - (D) OTHER INFORMATION: /label= Modified-site

- 42 -

/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 14
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 15
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 2
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Guanine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 3

- 43 -

(D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site
(B) LOCATION: 4
(D) OTHER INFORMATION: /label= Modified-site
/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site
(B) LOCATION: 5
(D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site
(B) LOCATION: 6
(D) OTHER INFORMATION: /label= Modified-site
/note= "Guanine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site
(B) LOCATION: 7
(D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site
(B) LOCATION: 8
(D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site
(B) LOCATION: 9
(D) OTHER INFORMATION: /label= Modified-site

- 44 -

/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 10
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 11
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 12
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 13
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 14
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 15
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

- 45 -

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1 5 10 15
- (2) INFORMATION FOR SEQ ID NO:10:
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: peptide
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 1
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Guanine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 2
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 3
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 4
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 5
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

- 46 -

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 6
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Guanine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 8
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Lys

1 5

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 2
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl

- 47 -

group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 3
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 4
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 6
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 8
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site

- 48 -

- (B) LOCATION: 9
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa

1 5

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 2
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 3
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 4
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- 49 -

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 6
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 8
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 9
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 10
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 11
- (D) OTHER INFORMATION: /label= Modified-site

- 50 -

/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 12
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 13
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 14
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 15
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 16
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids

- 51 -

- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: peptide
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 1
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 2
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 3
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 4
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 5
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 6
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl

- 52 -

group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 7

(D) OTHER INFORMATION: /label= Modified-site

/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 8

(D) OTHER INFORMATION: /label= Modified-site

/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 9

(D) OTHER INFORMATION: /label= Modified-site

/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 10

(D) OTHER INFORMATION: /label= Modified-site

/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 11

(D) OTHER INFORMATION: /label= Modified-site

/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 12

(D) OTHER INFORMATION: /label= Modified-site

/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 13

(D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(A) NAME/KEY: Modified-site
(B) LOCATION: 14
(D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(A) NAME/KEY: Modified-site
(B) LOCATION: 15
(D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(A) NAME/KEY: Modified-site
(B) LOCATION: 16
(D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1 5 10 15

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

```
(A) NAME/KEY: Modified-site  
(B) LOCATION: 1  
(D) OTHER INFORMATION: /label= Modified-site  
    /note= "Cytosine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."
```

- 54 -

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 2
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 3
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 4
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 6
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 8

- 55 -

- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 9
(D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 10
(D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1 5 10
- (2) INFORMATION FOR SEQ ID NO:15:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown
(ii) MOLECULE TYPE: peptide
(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 1
(D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 2
(D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
(A) NAME/KEY: Modified-site

- 56 -

- (B) LOCATION: 3
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 4
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 5
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 6
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 7
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 8
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
 - (A) NAME/KEY: Modified-site
 - (B) LOCATION: 9
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to

N-acetyl(2-aminoethyl)glycine through the N-acetyl group at position 1 of the heterocycle."

(A) NAME/KEY: Modified-site
(B) LOCATION: 10
(D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1 5 10

(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(A) NAME/KEY: Modified-site
(B) LOCATION: 1
(D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(A) NAME/KEY: Modified-site
(B) LOCATION: 2
(D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(A) NAME/KEY: Modified-site
(B) LOCATION: 3
(D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(A) NAME/KEY: Modified-site
(B) LOCATION: 4
(D) OTHER INFORMATION: /label= Modified-site

- 58 -

/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 6
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 8
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 9
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 10
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

- 59 -

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 11
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 12
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 13
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 14
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 15
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- 60 -

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 1

(D) OTHER INFORMATION: /label= Modified-site

/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 2

(D) OTHER INFORMATION: /label= Modified-site

/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 3

(D) OTHER INFORMATION: /label= Modified-site

/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 4

(D) OTHER INFORMATION: /label= Modified-site

/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 5

(D) OTHER INFORMATION: /label= Modified-site

/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 6

(D) OTHER INFORMATION: /label= Modified-site

/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

- 61 -

(B) LOCATION: 7

(D) OTHER INFORMATION: /label= Modified-site

/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 8

(D) OTHER INFORMATION: /label= Modified-site

/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 9

(D) OTHER INFORMATION: /label= Modified-site

/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 10

(D) OTHER INFORMATION: /label= Modified-site

/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 11

(D) OTHER INFORMATION: /label= Modified-site

/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 12

(D) OTHER INFORMATION: /label= Modified-site

/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 13

(D) OTHER INFORMATION: /label= Modified-site

/note= "Cytosine heterocycle is attached to

- 62 -

N-acetyl(2-aminoethyl)glycine through the N-acetyl group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 14
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to N-acetyl(2-aminoethyl)glycine through the N-acetyl group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 15
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to N-acetyl(2-aminoethyl)glycine through the N-acetyl group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 16
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to N-acetyl(2-aminoethyl)glycine through the N-acetyl group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 17
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to N-acetyl(2-aminoethyl)glycine through the N-acetyl group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 18
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to N-acetyl(2-aminoethyl)glycine through the N-acetyl group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 19
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to N-acetyl(2-aminoethyl)glycine through the N-acetyl group at position 1 of the heterocycle."

(ix) FEATURE:

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa |
| 1 | | | | 5 | | | | | 10 | | | | | 15 | |
| Xaa | Xaa | Xaa | Xaa | Lys | Lys | Lys | Lys | Lys | Lys | Lys | Lys | Lys | Lys | | |
| | | | | 20 | | | | | 25 | | | | | | |

(A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(A) NAME/KEY: Modified-site
(B) LOCATION: 1
(D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

```
(A) NAME/KEY: Modified-site
(B) LOCATION: 2
(D) OTHER INFORMATION: /label= Modified-site
    /note= "Thymine heterocycle is attached to
    N-acetyl(2-aminoethyl)glycine through the N-acetyl
    group at position 1 of the heterocycle."
```

```
(A) NAME/KEY: Modified-site  
(B) LOCATION: 3  
(D) OTHER INFORMATION: /label= Modified-site  
                        /note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."
```

(A) NAME/KEY: Modified-site
(B) LOCATION: 4
(D) OTHER INFORMATION: /label= Modified-site
/note= "Adenine heterocycle is attached to

- 64 -

N-acetyl(2-aminoethyl)glycine through the N-acetyl group at position 9 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 6
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl group at position 9 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 8
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 9
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 10
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl group at position 1 of the heterocycle."

(ix) FEATURE:

- 65 -

- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 11
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Guanine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 12
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 13
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 14
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 15
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 16
 - (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 17
 - (D) OTHER INFORMATION: /label= Modified-site

- 66 -

/note= "Guanine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 18
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 19
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 20
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa |
| 1 | | | | 5 | | | | | 10 | | | | | 15 | |
| Xaa | Xaa | Xaa | Xaa | Lys | | | | | | | | | | | |
| | | | | 20 | | | | | | | | | | | |

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /label= Modified-site

- 67 -

/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 2
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 3
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Guanine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 4
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 6
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /label= Modified-site

- 68 -

/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 8
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 9
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 10
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 11
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 12
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 13
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

- 69 -

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 14
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 15
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 16
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 17
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 18
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 19
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 20

(D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

[illegible]

N-acetyl(2-aminoethyl)glycine through the N-acetyl group at position 1 of the heterocycle."

N-acetyl(2-aminoethyl)glycine through the N-acetyl group at position 9 of the heterocycle."

N-acetyl(2-aminoethyl)glycine through the N-acetyl group at position 1 of the heterocycle."

N-acetyl(2-aminoethyl)glycine through the N-acetyl group at position 1 of the heterocycle."

- 71 -

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 6
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Guanine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 8
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 9
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 10
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 11

- 72 -

- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
- (B) LOCATION: 12
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
- (B) LOCATION: 13
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Guanine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
- (B) LOCATION: 14
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
- (B) LOCATION: 15
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
- (B) LOCATION: 16
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heteroside is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
- (B) LOCATION: 17
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl

- 73 -

group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 18
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 19
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of thr heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 20
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

| | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa |
| 1 | | | | 5 | | | | | 10 | | | | | 15 | | |
| Xaa | Xaa | Xaa | Xaa | Lys | | | | | | | | | | | | |
| | | | | 20 | | | | | | | | | | | | |

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 2

- 74 -

(D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site
(B) LOCATION: 3
(D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site
(B) LOCATION: 4
(D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site
(B) LOCATION: 5
(D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site
(B) LOCATION: 6
(D) OTHER INFORMATION: /label= Modified-site
/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site
(B) LOCATION: 7
(D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site
(B) LOCATION: 8
(D) OTHER INFORMATION: /label= Modified-site

- 75 -

/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 9

(D) OTHER INFORMATION: /label= Modified-site

/note= "Adenine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 9 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 10

(D) OTHER INFORMATION: /label= Modified-site

/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 11

(D) OTHER INFORMATION: /label= Modified-site

/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 12

(D) OTHER INFORMATION: /label= Modified-site

/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 13

(D) OTHER INFORMATION: /label= Modified-site

/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 14

(D) OTHER INFORMATION: /label= Modified-site

/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

- 76 -

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 15
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 16
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 17
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 18
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 19
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 20
- (D) OTHER INFORMATION: /label= Modified-site
/note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa

- 77 -

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
(B) LOCATION: 1
(D) OTHER INFORMATION: /label= Modified-site
 /note= "Cytosine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."

(ix) FEATURE:

- ```
(A) NAME/KEY: Modified-site
(B) LOCATION: 2
(D) OTHER INFORMATION: /label= Modified-site
 /note= "Thymine heterocycle is attached to
N-acetyl(2-aminoethyl)glycine through the N-acetyl
group at position 1 of the heterocycle."
```

(ix) FEATURE:

- (A) NAME/KEY: Modified-site  
(B) LOCATION: 3  
(D) OTHER INFORMATION: /label= Modified-site  
/note= "Guanine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 9 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site  
(B) LOCATION: 4  
(D) OTHER INFORMATION: /label= Modified-site  
/note= "Cytosine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site  
(B) LOCATION: 5  
(D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

- 78 -

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 6
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Guanine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 9 of the heterocycle."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Cytosine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 8
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Cytosine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 9
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 10
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Cytosine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 11
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 12



- 79 -

- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Guanine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 9 of the heterocycle."
- (ix) FEATURE:  
(A) NAME/KEY: Modified-site  
(B) LOCATION: 13  
(D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."
- (ix) FEATURE:  
(A) NAME/KEY: Modified-site  
(B) LOCATION: 14  
(D) OTHER INFORMATION: /label= Modified-site  
/note= "Cytosine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."
- (ix) FEATURE:  
(A) NAME/KEY: Modified-site  
(B) LOCATION: 15  
(D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."
- (ix) FEATURE:  
(A) NAME/KEY: Modified-site  
(B) LOCATION: 16  
(D) OTHER INFORMATION: /label= Modified-site  
/note= "Cytosine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."
- (ix) FEATURE:  
(A) NAME/KEY: Modified-site  
(B) LOCATION: 17  
(D) OTHER INFORMATION: /label= Modified-site  
/note= "Adenine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 9 of the heterocycle."
- (ix) FEATURE:  
(A) NAME/KEY: Modified-site  
(B) LOCATION: 18  
(D) OTHER INFORMATION: /label= Modified-site

- 80 -

/note= "Guanine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 9 of the heterocycle."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 19
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Guanine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 9 of the heterocycle."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 20
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
1 5 10 15  
Xaa Xaa Xaa Xaa Lys  
20

## (2) INFORMATION FOR SEQ ID NO:23:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: peptide

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 2
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

## (ix) FEATURE:

- 81 -

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 3
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 4
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Beta isoform of alanine."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 6
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Cytosine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 8
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 9
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl

- 82 -

group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 10
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 11
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Xaa Xaa Xaa Xaa Ala Xaa Xaa Xaa Xaa Xaa Xaa Lys  
1 5 10

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 2
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 3
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to

- 83 -

N-acetyl(2-aminoethyl)glycine through the N-acetyl group at position 1 of the heterocycle."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 4
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Beta isoform of alanine."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 6
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 8
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 9
- (D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 10

- 84 -

(D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

## (ix) FEATURE:

(A) NAME/KEY: Modified-site  
(B) LOCATION: 11  
(D) OTHER INFORMATION: /label= Modified-site  
/note= "Thymine heterocycle is attached to  
N-acetyl(2-aminoethyl)glycine through the N-acetyl  
group at position 1 of the heterocycle."

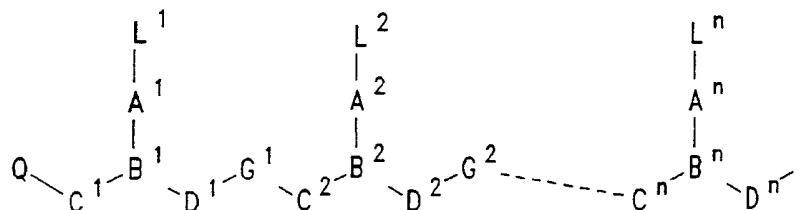
## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Xaa Xaa Xaa Xaa Ala Xaa Xaa Xaa Xaa Xaa Lys  
1                      5                      10

- 85 -

## WHAT IS CLAIMED IS:

1. A compound having the formula:



wherein:

n is at least 2,

5 each of  $L^1-L^n$  is independently selected from the group consisting of hydrogen, hydroxy,  $(C_1-C_4)$ alkanoyl, naturally occurring nucleobases, non-naturally occurring nucleobases, aromatic moieties, DNA intercalators, nucleobase-binding groups, heterocyclic moieties, and

10 reporter ligands, at least one of  $L^1-L^n$  being a naturally occurring nucleobase, a non-naturally occurring nucleobase, a DNA intercalator, or a nucleobase-binding group;

each of  $C^1-C^n$  is  $(CR^6R^7)_y$  where  $R^6$  is hydrogen and  $R^7$  is selected from the group consisting of the side chains of

15 naturally occurring alpha amino acids, or  $R^6$  and  $R^7$  are independently selected from the group consisting of hydrogen,  $(C_2-C_6)$ alkyl, aryl, aralkyl, heteroaryl, hydroxy,  $(C_1-C_6)$ alkoxy,  $(C_1-C_6)$ alkylthio,  $NR^3R^4$  and  $SR^5$ , where  $R^3$  and  $R^4$  are as defined above, and  $R^5$  is hydrogen,  $(C_1-C_6)$ alkyl, hydroxy-,

20 alkoxy-, or alkylthio- substituted  $(C_1-C_6)$ alkyl, or  $R^6$  and  $R^7$  taken together complete an alicyclic or heterocyclic system;

each of  $D^1-D^n$  is  $(CR^6R^7)_z$  where  $R^6$  and  $R^7$  are as defined above;

each of y and z is zero or an integer from 1 to 10,

25 the sum  $y + z$  being greater than 2 but not more than 10;

each of  $G^1-G^{n-1}$  is  $-NR^3CO-$ ,  $-NR^3CS-$ ,  $-NR^3SO-$  or  $-NR^3SO_2-$ , in either orientation, where  $R^3$  is as defined above;

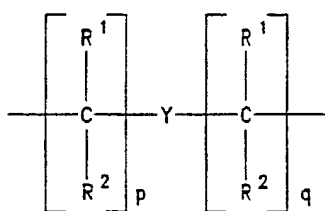
- 86 -

each of  $A^1-A^n$  and  $B^1-B^n$  are selected such that:

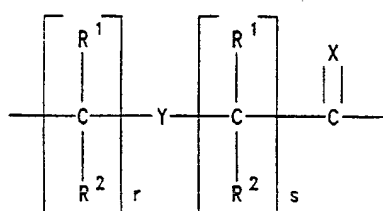
(a) A is a group of formula (IIa), (IIb) or (IIc), and B is N or  $R^3N^+$ , provided that at least one A is a group of formula (IIc); or

5 (b) A is a group of formula (IId) and B is CH; or

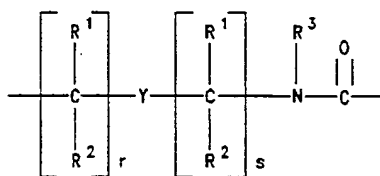
(c) A is a group of formula (IIa) or (IIb) and B is N or  $R^3N^+$ , provided at least one of y or z is not 1 or 2;



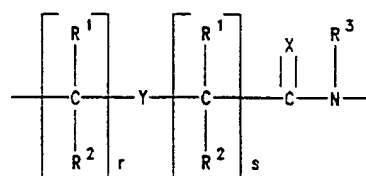
(IIa)



(IIb)



(IIc)



(IId)

10

where:

X is O, S, Se,  $NR^3$ ,  $CH_2$  or  $C(CH_3)_2$ ;

Y is a single bond, O, S or  $NR^4$ ;

15 each of p and q is zero or an integer from 1 to 5, the sum p+q being not more than 10;

each of r and s is zero or an integer from 1 to 5, the sum r+s being not more than 10;

20 each  $R^1$  and  $R^2$  is independently selected from the group consisting of hydrogen,  $(C_1-C_4)$ alkyl which may be hydroxy- or alkoxy- or alkylthio-substituted, hydroxy, alkoxy, alkylthio, amino and halogen; and

25 each  $R^3$  and  $R^4$  is independently selected from the group consisting of hydrogen,  $(C_1-C_4)$ alkyl, hydroxy- or alkoxy- or alkylthio-substituted  $(C_1-C_4)$ alkyl, hydroxy, alkoxy, alkylthio and amino;



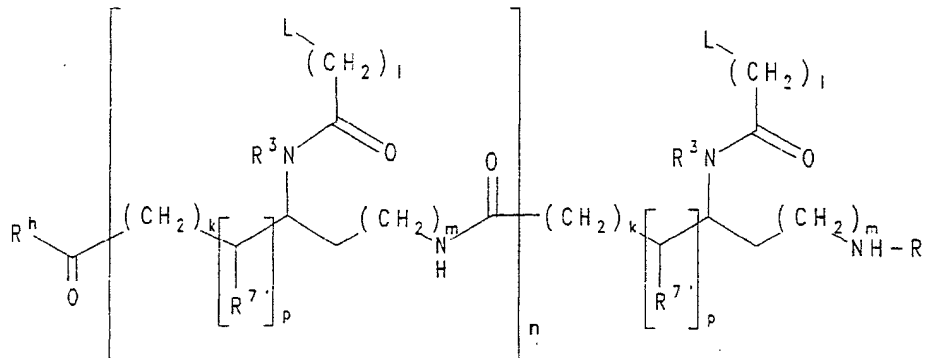
- 87 -

Q is  $-\text{CO}_2\text{H}$ ,  $-\text{CONR}'\text{R}''$ ,  $-\text{SO}_3\text{H}$  or  $-\text{SO}_2\text{NR}'\text{R}''$  or an activated derivative of  $-\text{CO}_2\text{H}$  or  $-\text{SO}_3\text{H}$ ; and

I is  $-\text{NHR}'''\text{R}''''$  or  $-\text{NR}'''\text{C}(\text{O})\text{R}''''$ , where  $\text{R}'$ ,  $\text{R}''$ ,  $\text{R}'''$  and  $\text{R}''''$  are independently selected from the group  
5 consisting of hydrogen, alkyl, amino protecting groups, reporter ligands, intercalators, chelators, peptides, proteins, carbohydrates, lipids, steroids, oligonucleotides and soluble and non-soluble polymers.

- 88 -

2. The compound of claim 1 having the formula:



wherein:

each L is independently selected from the group consisting of hydrogen, phenyl, heterocyclic moieties, naturally occurring nucleobases, and non-naturally occurring nucleobases;

each R<sup>7'</sup> is independently selected from the group consisting of hydrogen and the side chains of naturally occurring alpha amino acids;

n is an integer from 1 to 60,

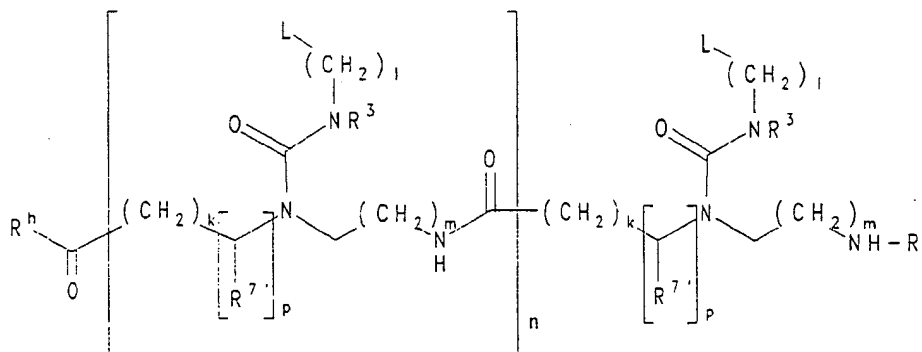
each k, l, and m is, independently, zero or an integer from 1 to 5;

each p is zero or 1;

R<sup>h</sup> is OH, NH<sub>2</sub> or -NHLysNH<sub>2</sub>; and

R<sup>i</sup> is H or COCH<sub>3</sub>.

3. The compound of claim 1 having the formula:



- 89 -

wherein:

each L is independently selected from the group consisting of hydrogen, phenyl, heterocyclic moieties, naturally occurring nucleobases, and non-naturally occurring nucleobases;

each R<sup>7</sup> is independently selected from the group consisting of hydrogen and the side chains of naturally occurring alpha amino acids;

n is an integer from 1 to 60,

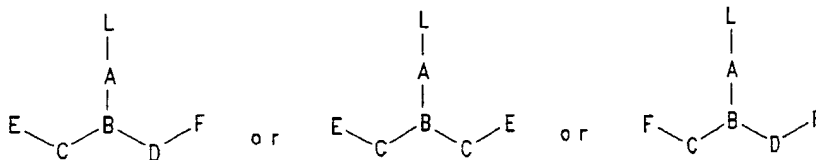
each k, l, and m is, independently, zero or an integer from 1 to 5;

each p is zero or 1;

R<sup>h</sup> is OH, NH<sub>2</sub> or -NHLysNH<sub>2</sub>; and

R<sup>i</sup> is H or COCH<sub>3</sub>.

4. A compound having one of the following formulas:



wherein:

L is selected from the group consisting of hydrogen, hydroxy, (C<sub>1</sub>-C<sub>4</sub>)alkanoyl, naturally occurring nucleobases, non-naturally occurring nucleobases, aromatic moieties, DNA intercalators, nucleobase-binding groups, and heterocyclic moieties, reporter ligands, wherein amino groups are, optionally, protected by amino protecting groups;

each C is (CR<sup>6</sup>R<sup>7</sup>)<sub>y</sub>, where R<sup>6</sup> is hydrogen and R<sup>7</sup> is selected from the group consisting of the side chains of naturally occurring alpha amino acids, or R<sup>6</sup> and R<sup>7</sup> are independently selected from the group consisting of hydrogen, (C<sub>2</sub>-C<sub>6</sub>)alkyl, aryl, aralkyl, heteroaryl, hydroxy, (C<sub>1</sub>-C<sub>6</sub>)alkoxy, (C<sub>1</sub>-C<sub>6</sub>)alkylthio, NR<sup>3</sup>R<sup>4</sup> and SR<sup>5</sup>, where R<sup>3</sup> and R<sup>4</sup> are as defined above, and R<sup>5</sup> is hydrogen, (C<sub>1</sub>-C<sub>6</sub>)alkyl, hydroxy-,

- 90 -

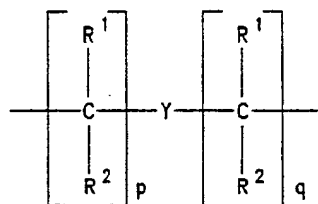
alkoxy-, or alkylthio- substituted (C<sub>1</sub>-C<sub>6</sub>)alkyl, or R<sup>6</sup> and R<sup>7</sup> taken together complete an alicyclic or heterocyclic system;

each D is (CR<sup>6</sup>R<sup>7</sup>)<sub>z</sub> where R<sup>6</sup> and R<sup>7</sup> are as defined above;

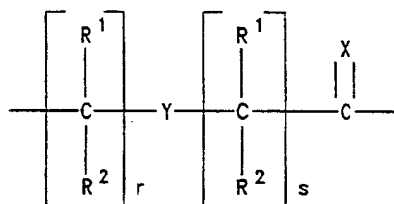
5 each of y and z is zero or an integer from 1 to 10, the sum y + z being greater than 2 but not more than 10;

A and B are selected such that:

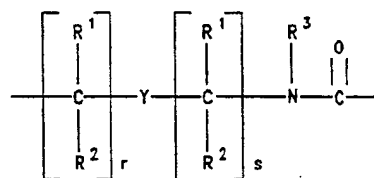
- (a) A is a group of formula (IIc) and B is N or R<sup>3</sup>N<sup>+</sup>; or
- 10 (b) A is a group of formula (IId) and B is CH; or
- (c) A is a group of formula (IIa) or (IIb) and B is N or R<sup>3</sup>N<sup>+</sup>, provided at least one of y or z is not 1 or 2;



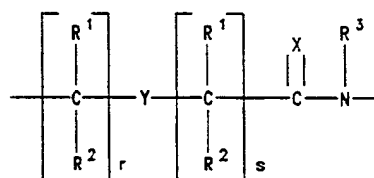
(IIa)



(IIb)



(IIc)



(IId)

where:

X is O, S, Se, NR<sup>3</sup>, CH<sub>2</sub> or C(CH<sub>3</sub>)<sub>2</sub>;

Y is a single bond, O, S or NR<sup>4</sup>;

20 each of p and q is zero or an integer from 1 to 5, the sum p+q being not more than 10;

each of r and s is zero or an integer from 1 to 5, the sum r+s being not more than 10;

25 each R<sup>1</sup> and R<sup>2</sup> is independently selected from the group consisting of hydrogen, (C<sub>1</sub>-C<sub>4</sub>)alkyl which may be hydroxy- or alkoxy- or alkylthio-

- 91 -

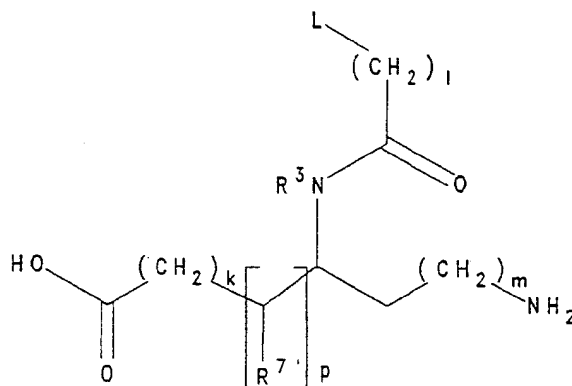
substituted, hydroxy, alkoxy, alkylthio, amino and halogen; and

each  $R^3$  and  $R^4$  is independently selected from the group consisting of hydrogen,  $(C_1-C_4)$ alkyl, hydroxy- or alkoxy- or alkylthio-substituted  $(C_1-C_4)$ alkyl, hydroxy, alkoxy, alkylthio and amino;

each E is COOH, CSOH, SOOH, SO<sub>2</sub>OH or an activated or protected derivative thereof; and

each F is  $NHR^3$  or  $NPgR^3$ , where  $R^3$  is as defined above, and Pg is an amino protecting group.

5. The compound of claim 4 having the formula:



wherein:

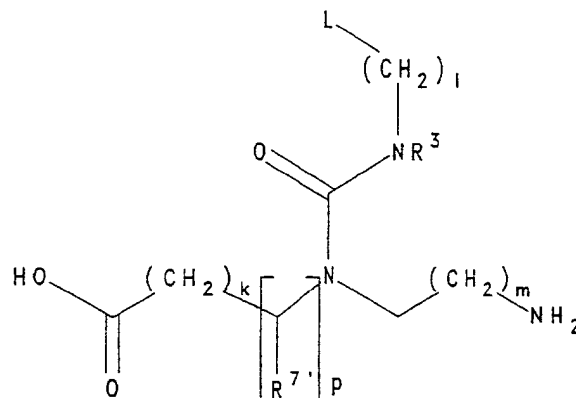
each L is independently selected from the group consisting of hydrogen, phenyl, heterocyclic moieties, naturally occurring nucleobases, and non-naturally occurring nucleobases;

each  $R^{7'}$  is independently selected from the group consisting of hydrogen and the side chains of naturally occurring alpha amino acids; and

each k, l, and m is, independently, zero or an integer from 1 to 5.

- 92 -

6. The compound of claim 4 having the formula:



wherein:

each L is independently selected from the group consisting of hydrogen, phenyl, heterocyclic moieties, naturally occurring nucleobases, and non-naturally occurring nucleobases;

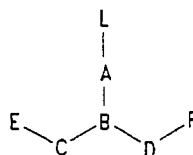
each R<sup>7'</sup> is independently selected from the group consisting of hydrogen and the side chains of naturally occurring alpha amino acids; and

each k, l, and m is, independently, zero or an integer from 1 to 5.

7. A process for preparing a compound according to claim 1, comprising the steps of:

A) providing a polymer substrate, said polymer being functionalized with a chemical group capable of forming an anchoring linkage with an amino acid;

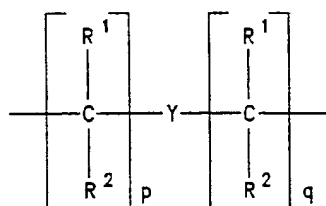
B) coupling said polymer with a first amino acid through said anchoring linkage, said first amino acid having formula (IV):



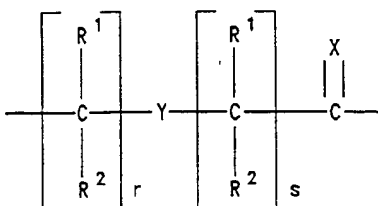
(IV)

wherein:

- L is selected from the group consisting of naturally occurring nucleobases, non-naturally occurring nucleobases, aromatic moieties, DNA intercalators, nucleobase-binding groups, heterocyclic moieties, and reporter ligands, wherein amino groups are, optionally, protected by amino protecting groups;
- each C is  $(CR^6R^7)_y$  where  $R^6$  is hydrogen and  $R^7$  is selected from the group consisting of the side chains of naturally occurring alpha amino acids, or  $R^6$  and  $R^7$  are independently selected from the group consisting of hydrogen,  $(C_2-C_6)$ alkyl, aryl, aralkyl, heteroaryl, hydroxy,  $(C_1-C_6)$ alkoxy,  $(C_1-C_6)$ alkylthio,  $NR^3R^4$  and  $SR^5$ , where  $R^3$  and  $R^4$  are as defined above, and  $R^5$  is hydrogen,  $(C_1-C_6)$ alkyl, hydroxy-, alkoxy-, or alkylthio- substituted  $(C_1-C_6)$ alkyl, or  $R^6$  and  $R^7$  taken together complete an alicyclic or heterocyclic system;
- each D is  $(CR^6R^7)_z$  where  $R^6$  and  $R^7$  are as defined above;
- each of y and z is zero or an integer from 1 to 10, the sum  $y + z$  being greater than 2 but not more than 10;
- A and B are selected such that:
- (a) A is a group of formula (IIc) and B is N or  $R^3N^+$ ; or
- (b) A is a group of formula (IIId) and B is CH; or
- (c) A is a group of formula (IIa) or (IIb) and B is N or  $R^3N^+$ , provided at least one of y or z is not 1 or 2;

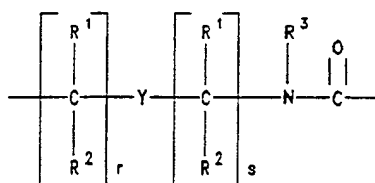


(IIa)

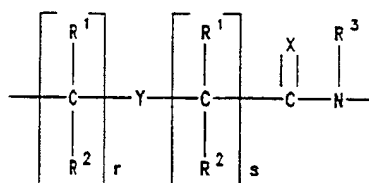


(IIb)

- 94 -



(IIc)



(IIId)

where:

X is O, S, Se, NR<sup>3</sup>, CH<sub>2</sub> or C(CH<sub>3</sub>)<sub>2</sub>;

Y is a single bond, O, S or NR<sup>4</sup>;

5 each of p and q is zero or an integer from 1 to 5, the sum p+q being not more than 10;

each of r and s is zero or an integer from 1 to 5, the sum r+s being not more than 10;

10 each R<sup>1</sup> and R<sup>2</sup> is independently selected from the group consisting of hydrogen, (C<sub>1</sub>-C<sub>4</sub>)alkyl which may be hydroxy- or alkoxy- or alkylthio-substituted, hydroxy, alkoxy, alkylthio, amino and halogen; and

15 each R<sup>3</sup> and R<sup>4</sup> is independently selected from the group consisting of hydrogen, (C<sub>1</sub>-C<sub>4</sub>)alkyl, hydroxy- or alkoxy- or alkylthio-substituted (C<sub>1</sub>-C<sub>4</sub>)alkyl, hydroxy, alkoxy, alkylthio and amino;

each E is COOH, CSOH, SOOH, SO<sub>2</sub>OH or an activated or protected derivative thereof; and

20 each F is NHR<sup>3</sup> or NPgR<sup>3</sup>, where R<sup>3</sup> is as defined above, and Pg is an amino protecting group;

C) removing said amino protecting group from said coupled first amino acid to generate a free amino group; and

25 D) reacting said free amino group with a second amino acid having formula (IV) to form a peptide chain.

8. The process of claim 7 further comprising the steps of:

E) removing said amino protecting group from said second amino acid to generate a terminal free amino group on  
30 said peptide chain; and



- 95 -

F) reacting said free amino group on said peptide chain with a further amino acid having formula (IV) to lengthen said peptide chain.

9. The process of claim 8 wherein steps E and F  
5 are performed a plurality of times.

10. The process of claim 8 further comprising removing at least one protecting group remaining on the amino acid moieties of the peptide chain.

11. The process of claim 7 further comprising  
10 cleaving said anchoring linkage without substantially degrading said peptide chain.

12. The process of claim 7 wherein the polymer substrate contains polystyrene, polyacrylamide, silica, a composite material, cotton, or a derivative thereof.

15 13. The process of claim 8 wherein the chemical group capable of forming said anchoring linkage is chloro-, bromo- and iodo-substituted alkyl, amino-substituted alkyl, amino and aryl-substituted alkyl, amino- and alkylaryl-substituted alkyl, hydroxy-substituted alkyl, or a derivative  
20 thereof having a spacer group that can be cleaved substantially without degradation of said polypeptide.

14. The process of claim 13 wherein chloro-substituted alkyl is chloromethyl, amino-substituted alkyl is aminomethyl, amino- and alkyl-substituted aryl is  $\alpha$ -  
25 aminobenzyl, amino- and alkylaryl-substituted alkyl is selected from the group consisting of  $\alpha$ -amino-3- and  $\alpha$ -amino-4-methylbenzyl, and hydroxy-substituted alkyl is hydroxymethyl.

- 96 -

15. The process of claim 13 wherein:  
the chemical group is derived from an amino-containing moiety  
selected from amino-substituted alkyl, amino- and aryl  
substituted alkyl, and amino- and alkylaryl-substituted  
5 alkyl; and

the chemical group includes a spacer group derived  
from the group consisting of 4-(haloalkyl)aryl-lower alkanolic  
acids, Boc-aminoacyl-4-(oxymethyl)aryl-lower alkanolic acids,  
N-Boc-p-acylbenzhydrylamines, N-Boc-4'-(lower alkyl)-p-  
10 acylbenzhydrylamines, N-Boc-4'-(lower alkoxy)-p-  
acylbenzhydrylamines, and 4-hydroxymethylphenoxy-lower  
alkanoic acids.

16. A process for sequence-specific recognition of  
a double-stranded polynucleotide, comprising contacting said  
15 polynucleotide with a compound that is different from natural  
RNA and that binds to one strand of the polynucleotide,  
thereby displacing the other strand, said being the compound  
of claim 1.

17. A process for modulating the expression of a  
20 gene in an organism, comprising administering to said  
organism a compound according to claim 1 that specifically  
binds to DNA or RNA deriving from said gene, said compound  
being the compound of claim 1.

18. The process of claim 17 wherein said  
25 modulation includes inhibiting transcription of said gene.

19. The process of claim 17 wherein said  
modulation includes inhibiting replication of said gene.

20. A process for treating conditions associated  
with undesired protein production in an organism, comprising  
30 contacting said organism with an effective amount of a  
compound according to claim 1 that specifically binds with

- 97 -

DNA or RNA deriving from a gene controlling said protein production.

21. A process for inducing degradation of DNA or RNA in cells of an organism, comprising administering to said  
5 organism a compound according to claim 1 that specifically binds to said DNA or RNA.

22. A process for killing cells or virus, comprising contacting said cells or virus with a compound according to claim 1 that specifically binds to a portion of  
10 the genome of said cells or virus.

23. A pharmaceutical composition comprising a compound according to claim 1 and at least one pharmaceutically effective carrier, binder, thickener, diluent, buffer, preservative, or surface active agent.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                               |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|
| (51) International Patent Classification <sup>5</sup> :<br>C07H 21/00, C12Q 1/68, C07K 15/00                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |  | A3                                                                                                                                                                                                                                                                                                                                                                                                                                                       | (11) International Publication Number: <b>WO 94/25477</b><br>(43) International Publication Date: 10 November 1994 (10.11.94) |
| (21) International Application Number: PCT/IB94/00142<br>(22) International Filing Date: 25 April 1994 (25.04.94)<br>(30) Priority Data:<br>08/054,363 26 April 1993 (26.04.93) US<br>(71)(72) Applicants and Inventors: NIELSEN, Peter, E. [DK/DK];<br>Hjortevanget 509, DK-2980 Kokkedal (DK). BUCHARDT,<br>Ole [DK/DK]; Søndergardsvej 73, DK-3500 Værløse (DK).<br>EGHOLM, Michael [DK/DK]; Johnstrup Alle 3, DK-1923<br>Frederiksberg (DK). BERG, Rolf, H. [DK/DK]; Strand-<br>vænget 6, DK-2960 Rungsted Kyst (DK).<br>(74) Agent: HALLYBONE, Huw, George; Carpmals & Ransford,<br>43 Bloomsbury Square, London WC1A 2RA (GB). |  | (81) Designated States: AU, BR, CA, FI, HU, JP, KR, NO,<br>European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR,<br>IE, IT, LU, MC, NL, PT, SE).<br><br>Published<br><i>With international search report.<br/>Before the expiration of the time limit for amending the<br/>claims and to be republished in the event of the receipt of<br/>amendments.</i><br>(88) Date of publication of the international search report:<br>22 December 1994 (22.12.94) |                                                                                                                               |
| (54) Title: NOVEL PEPTIDE NUCLEIC ACIDS                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                               |
| (57) Abstract<br><br>A novel class of compounds, known as peptide nucleic acids, bind complementary ssDNA and RNA strands more strongly than a corresponding DNA. The peptide nucleic acids generally comprise ligands such as naturally occurring DNA bases attached to a peptide backbone through a suitable linker.                                                                                                                                                                                                                                                                                                               |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                               |

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

|    |                          |    |                                       |    |                          |
|----|--------------------------|----|---------------------------------------|----|--------------------------|
| AT | Austria                  | GB | United Kingdom                        | MR | Mauritania               |
| AU | Australia                | GE | Georgia                               | MW | Malawi                   |
| BB | Barbados                 | GN | Guinea                                | NE | Niger                    |
| BE | Belgium                  | GR | Greece                                | NL | Netherlands              |
| BF | Burkina Faso             | HU | Hungary                               | NO | Norway                   |
| BG | Bulgaria                 | IE | Ireland                               | NZ | New Zealand              |
| BJ | Benin                    | IT | Italy                                 | PL | Poland                   |
| BR | Brazil                   | JP | Japan                                 | PT | Portugal                 |
| BY | Belarus                  | KE | Kenya                                 | RO | Romania                  |
| CA | Canada                   | KG | Kyrgyzstan                            | RU | Russian Federation       |
| CF | Central African Republic | KP | Democratic People's Republic of Korea | SD | Sudan                    |
| CG | Congo                    | KR | Republic of Korea                     | SE | Sweden                   |
| CH | Switzerland              | KZ | Kazakhstan                            | SI | Slovenia                 |
| CI | Côte d'Ivoire            | LI | Liechtenstein                         | SK | Slovakia                 |
| CM | Cameroon                 | LK | Sri Lanka                             | SN | Senegal                  |
| CN | China                    | LU | Luxembourg                            | TD | Chad                     |
| CS | Czechoslovakia           | LV | Latvia                                | TC | Togo                     |
| CZ | Czech Republic           | MC | Monaco                                | TJ | Tajikistan               |
| DE | Germany                  | MD | Republic of Moldova                   | TT | Trinidad and Tobago      |
| DK | Denmark                  | MG | Madagascar                            | UA | Ukraine                  |
| ES | Spain                    | ML | Mali                                  | US | United States of America |
| FI | Finland                  | MN | Mongolia                              | UZ | Uzbekistan               |
| FR | France                   |    |                                       | VN | Viet Nam                 |
| GA | Gabon                    |    |                                       |    |                          |

## INTERNATIONAL SEARCH REPORT

International Application No.

PCT/IB 94/00142

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 5 C07H21/00 C12Q1/68 C07K15/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 5 C07H C12Q C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                               | Relevant to claim No. |
|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| X          | WO,A,92 20703 (O BUCHARDT ET AL.) 26<br>November 1992<br>see the whole document<br>---                                                                                                                                                                           | 1,3,4,<br>6-23        |
| X          | WO,A,92 20702 (O BUCHARDT ET AL.) 26<br>November 1992<br>see the whole document<br>---                                                                                                                                                                           | 1,3,4,<br>6-23        |
| X          | SCIENCE.,<br>vol.254, no.5037, 6 December 1991,<br>LANCASTER, PA US<br>pages 1497 - 1500<br>P E NIELSEN ET AL 'Sequence-selective<br>recognition of DNA by strand displacement<br>with a thymine-substituted polyamide'<br>see the whole document<br>---<br>-/-- | 1,3,4,<br>6-23        |

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

## \* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

21 September 1994

Date of mailing of the international search report

27. 10. 94

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax (+31-70) 340-3016

Authorized officer

Masturzo, P

## INTERNATIONAL SEARCH REPORT

International Application No.

PCT/IB 94/00142

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                          | Relevant to claim No. |
|----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| X        | PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA.,<br>vol.90, no.5, 1 March 1993, WASHINGTON US<br>pages 1667 - 1670<br>D Y CHERNY ET AL. 'DNA unwinding upon strand-displacement binding of a thymine-substituted polyamide to double-stranded DNA'<br>see the whole document<br>--- | 1,3,4,<br>6-23        |
| X        | JOURNAL OF THE AMERICAN CHEMICAL SOCIETY.,<br>vol.114, no.24, 18 November 1992, GASTON, PA US<br>pages 9677 - 9678<br>M EGHOLM ET AL. 'Recognition of guanine and adenine in DNA by cytosine and thymine containing peptide nucleic acid (PNA)'<br>see the whole document<br>---            | 1,3,4,<br>6-23        |
| A        | WO,A,86 05518 (J SUMMERTON ET AL.) 25<br>September 1986<br>see the whole document<br>-----                                                                                                                                                                                                  | 1-23                  |

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/IB 94/00142

**Box I** Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority; namely:  
Remark: Even though claims 20-21 and 17-19, 22 (partially) refer to a method of treatment of the human body, the search was carried out and based on the alleged effect of the products.
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:  
Since the general formula of claims 1 & 4 have no fixed invariable element, the search was restricted to the substructures of claims 2-3 and 5-6, process for their preparation etc., provided that the group L represents a nucleobase.
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box II** Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.  
☐ No protest accompanied the payment of additional search fees.



## INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No.

PCT/IB 94/00142

| Patent document<br>cited in search report | Publication<br>date | Patent family<br>member(s) |          | Publication<br>date |
|-------------------------------------------|---------------------|----------------------------|----------|---------------------|
| WO-A-9220703                              | 26-11-92            | AU-A-                      | 1880692  | 30-12-92            |
|                                           |                     | AU-A-                      | 1884392  | 30-12-92            |
|                                           |                     | CA-A-                      | 2109320  | 25-11-92            |
|                                           |                     | WO-A-                      | 9220702  | 26-11-92            |
|                                           |                     | EP-A-                      | 0586474  | 16-03-94            |
|                                           |                     | EP-A-                      | 0586618  | 16-03-94            |
| -----                                     |                     |                            |          |                     |
| WO-A-9220702                              | 26-11-92            | AU-A-                      | 1880692  | 30-12-92            |
|                                           |                     | AU-A-                      | 1884392  | 30-12-92            |
|                                           |                     | CA-A-                      | 2109320  | 25-11-92            |
|                                           |                     | WO-A-                      | 9220703  | 26-11-92            |
|                                           |                     | EP-A-                      | 0586474  | 16-03-94            |
|                                           |                     | EP-A-                      | 0586618  | 16-03-94            |
| -----                                     |                     |                            |          |                     |
| WO-A-8605518                              | 25-09-86            | AU-A-                      | 5661386  | 13-10-86            |
|                                           |                     | AU-A-                      | 5698186  | 13-10-86            |
|                                           |                     | CA-A-                      | 1268404  | 01-05-90            |
|                                           |                     | DE-A-                      | 3687030  | 03-12-92            |
|                                           |                     | EP-A, B                    | 0216860  | 08-04-87            |
|                                           |                     | EP-A-                      | 0215942  | 01-04-87            |
|                                           |                     | JP-T-                      | 62502338 | 10-09-87            |
|                                           |                     | JP-T-                      | 62502357 | 10-09-87            |
|                                           |                     | WO-A-                      | 8605519  | 25-09-86            |
|                                           |                     | US-A-                      | 5142047  | 25-08-92            |
|                                           |                     | US-A-                      | 5034506  | 23-07-91            |
|                                           |                     | US-A-                      | 5235033  | 10-08-93            |
|                                           |                     | US-A-                      | 5185444  | 09-02-93            |
|                                           |                     | US-A-                      | 5217866  | 08-06-93            |
| -----                                     |                     |                            |          |                     |